

On the tail behaviour of the distribution function of the maximum for the partial sums of a class of i.i.d. random variables.

Péter Major

Alfréd Rényi Mathematical Institute of the Hungarian Academy of Science

e-mail address: major.peter@renyi.mta.hu

Summary. We take an L_1 -dense class of functions \mathcal{F} on a measurable space (X, \mathcal{X}) and a sequence of i.i.d. X -valued random variables ξ_1, \dots, ξ_n , and give a good estimate on the tail behaviour of $\sup_{f \in \mathcal{F}} \sum_{j=1}^n f(\xi_j)$ if the conditions $\sup_{x \in X} |f(x)| \leq 1$, $Ef(\xi_1) = 0$ and $Ef(\xi_1)^2 < \sigma^2$ with some $0 \leq \sigma \leq 1$ hold for all $f \in \mathcal{F}$. Roughly speaking this estimate states that under some natural conditions the above considered supremum is not much larger than the worst element taking part in it. The proof heavily depends on the main result of paper [3]. Here we have to deal with such a problem where the classical methods worked out to investigate the behaviour of Gaussian or almost Gaussian random variables do not work.

1. Introduction.

The main result of this paper is an estimate about the tail-distribution of the supremum of partial sums of i.i.d. random variables presented in Theorem 1 together with an extension of it that provides an estimate for this tail-distribution in some cases not covered in Theorem 1. At first glance these results may look rather complicated, but as I try to explain in Section 2 they yield sharp estimates under natural conditions. They express such a fact that under some natural conditions we can get an almost as good bound for the supremum of an appropriately defined class of partial sums as for one term taking part in this supremum. Before presenting these results I recall the definition of L_1 -dense classes of functions, a notion that appears in the formulation of Theorem 1.

Definition of L_1 -dense classes of functions. Let a measurable space (X, \mathcal{X}) be given together with a class of \mathcal{X} measurable, real valued functions \mathcal{F} on this space. The class of functions \mathcal{F} is called an L_1 -dense class of functions with parameter D and exponent L if for all numbers $0 < \varepsilon \leq 1$ and probability measures ν on the space (X, \mathcal{X}) there exists a finite ε -dense subset $\mathcal{F}_{\varepsilon, \nu} = \{f_1, \dots, f_m\} \subset \mathcal{F}$ in the space $L_1(X, \mathcal{X}, \nu)$ with $m \leq D\varepsilon^{-L}$ elements, i.e. there exists such a set $\mathcal{F}_{\varepsilon, \nu} \subset \mathcal{F}$ with $m \leq D\varepsilon^{-L}$ elements for which $\inf_{f_j \in \mathcal{F}_{\varepsilon, \nu}} \int |f - f_j| d\nu < \varepsilon$ for all functions $f \in \mathcal{F}$.

Theorem 1 yields the following estimate.

Theorem 1. Let a sequence of independent, identically distributed random variables ξ_1, \dots, ξ_n , $n \geq 2$, with values in a measurable space (X, \mathcal{X}) and with some distribution μ be given together with a countable L_1 -dense class of functions \mathcal{F} with parameter $D \geq 1$ and exponent $L \geq 1$ on the space (X, \mathcal{X}) such that $\sup_{x \in X} |f(x)| \leq 1$, $\int f(x)\mu(dx) = 0$, and $\int f^2(x)\mu(dx) \leq \sigma^2$ with some number $0 \leq \sigma^2 \leq 1$ for all $f \in \mathcal{F}$. Define the normalized

random sums $S_n(f) = \frac{1}{\sqrt{n}} \sum_{j=1}^n f(\xi_j)$ for all $f \in \mathcal{F}$. There are some universal constants $C_j > 0$, $1 \leq j \leq 5$, (such that also the inequality $C_2 < 1$ holds), for which the inequality

$$P \left(\sup_{f \in \mathcal{F}} |S_n(f)| \geq v \right) \leq C_1 e^{-C_2 \sqrt{n} v \log(v/\sqrt{n}\sigma^2)} \quad \text{for all } v \geq u(\sigma) \quad (1.1)$$

holds if one of the following conditions is satisfied.

- (a) $\sigma^2 \leq \frac{1}{n^{200}}$, and $u(\sigma) = \frac{C_3}{\sqrt{n}}(L + \frac{\log D}{\log n})$,
- (b) $\frac{1}{n^{200}} < \sigma^2 \leq \frac{\log n}{8n}$, and $u(\sigma) = \frac{C_4}{\sqrt{n}} \left(L \frac{\log n}{\log(\frac{\log n}{n\sigma^2})} + \log D \right)$,
- (c) $\frac{\log n}{8n} < \sigma^2 \leq 1$, and $u(\sigma) = \frac{C_5}{\sqrt{n}}(n\sigma^2 + L \log n + \log D)$.

I complete the result of Theorem 1 with an extension which is actually a repetition of Theorem 4.1 in [2]. It yields an estimate for $P \left(\sup_{f \in \mathcal{F}} |S_n(f)| \geq v \right)$ in cases not covered in Theorem 1.

Extension of Theorem 1. *Let us consider, similarly to Theorem 1, a sequence of independent, identically distributed random variables ξ_1, \dots, ξ_n , $n \geq 2$, with values in a measurable space (X, \mathcal{X}) with some distribution μ together with a countable L_1 -dense class of functions \mathcal{F} with parameter $D \geq 1$ and exponent $L \geq 1$ on the space (X, \mathcal{X}) such that $\sup_{x \in X} |f(x)| \leq 1$, $\int f(x)\mu(dx) = 0$, and $\int f^2(x)\mu(dx) \leq \sigma^2$ with some number $0 \leq \sigma^2 \leq 1$ for all $f \in \mathcal{F}$. The supremum of the normalized partial sums $S_n(f)$, $f \in \mathcal{F}$, introduced in Theorem 1 satisfies the inequality*

$$P \left(\sup_{f \in \mathcal{F}} |S_n(f)| \geq v \right) \leq C \exp \left\{ -\alpha \frac{v^2}{\sigma^2} \right\} \quad (1.2)$$

with appropriate (universal) constants $\alpha > 0$, $C > 0$ and $C_6 > 0$ if $\frac{\log n}{8n} < \sigma^2 \leq 1$, $\bar{u}(\sigma) \leq v \leq \sqrt{n}\sigma^2$, where $\bar{u}(\sigma)$ is defined as $\bar{u}(\sigma) = C_6\sigma(L^{3/4} \log^{1/2} \frac{2}{\sigma} + (\log D)^{3/4})$.

The value $\frac{\log n}{8n}$ determining the boundary between cases (b) and (c) in Theorem 1 could be replaced by $\alpha \frac{\log n}{n}$ with any number $0 < \alpha < 1$. To see this one has to check that the formula defining $u(\sigma)$ in cases (b) and (c) give a value of the same order if $\sigma^2 \sim \alpha \frac{\log n}{n}$ with $0 < \alpha < 1$. I chose the parameter $\alpha = \frac{1}{8}$ because some calculations were simpler with such a choice. Let me remark that a similar statement holds for the value of boundary n^{-200} between cases (a) and (b). This could have been replaced by $n^{-\beta}$ with any $\beta > 1$.

In Section 2 I try to explain why the above results are natural, in Section 3 I present their proof, and in Section 4 I make some additional remarks. I finish this section with

a short comparison of the results of this paper with some similar results of Talagrand in [6].

In both works the magnitude of the supremum of partial sums of i.i.d. random variables are studied, and behind the results there is their implicit comparison with analogous estimates about the supremum of Gaussian random variables.

The analogous problems about the supremum of Gaussian random variables can be well investigated by the so-called chaining argument, whose best, sharpest version is worked out in [6]. The estimates about the supremum of Gaussian random variables can be simply generalized for the supremum of other classes random variables if the tail distributions of the differences of the elements from these classes of random variables satisfy an estimate similar to the corresponding estimate in the Gaussian case. If we consider partial sums of independent random variables, then the tail distributions of these partial sums satisfy only a weaker estimate. Hence some additional conditions have to be imposed in order to get good results. Both here and in [6] good estimates are given for the tail distribution of partial sums of i.i.d. random variables under some additional conditions. But these additional conditions are different in the two works, and in my opinion the difference between them is not such a technical detail as it may seem at first sight.

Talagrand extends the chaining argument to other models by exploiting that under some additional conditions a better (Gaussian) estimate can be given for the tail distribution of sums i.i.d. random variables. His proof can be considered as the extension of a Gaussian argument to a more general class of models with ‘almost Gaussian behaviour’. The additional condition of this paper about the existence of an L_1 -dense class of functions has a different character. It is useful to guarantee that the influence of some unpleasant ‘non-Gaussian effects’ in the model we are working with is small. The proof of this fact demands an argument different from the usual methods applied in the Gaussian case. I do not write down the details about the difference of the two methods, because I did it in Chapter 18 of [2] at pp. 235–237. Let me remark that the results obtained with their help cannot substitute each other. There are problems where the first one is useful and there are problems where the second one.

2. Discussion on the conditions of these results.

Our goal was to give sharp estimate for the supremum of a class of normalized partial sums $S_n(f)$ defined in Theorem 1 if the functions f are elements of an L_1 -dense class of functions \mathcal{F} that satisfies the conditions of Theorem 1. We have to explain why formulas (1.1) and (1.2) provide the right estimate in this problem, and why we had to impose the conditions $v \geq u(\sigma)$ and $\bar{u}(\sigma) \leq v \leq \sqrt{n}\sigma^2$ in them. We prove such estimates which depend on some universal multiplying constants whose optimal choice we do not investigate. Besides, we try to give a good value for the functions $u(\sigma)$ and $\bar{u}(\sigma)$ only in the case when the parameter D and exponent L of the L_1 -dense class of functions \mathcal{F} are bounded by a fixed number not depending on the parameter σ^2 . If the parameter D or exponent L is very large, then a different function $u(\sigma)$ could be chosen that provides a sharper result.

If we disregard the value of the universal constants appearing in our estimates then we can say that the estimate (1.1) for the tail distribution of the supremum we consider and the estimate of Bennett's inequality for the tail distribution of a single term in this supremum agree, at least in the case if we consider the estimate of Bennett's inequality at level $v \geq 2\sqrt{n}\sigma^2$. (This follows e.g. from formula (3.3) in this paper. We recalled Bennett's inequality, and formula (3.3) is a part of it.) On the other hand, we considered in Theorem 1 only such levels v where this condition is satisfied, since $u(\sigma) \geq 2\sqrt{n}\sigma$ in all cases of Theorem 1. Moreover, there are examples that show that inequality (3.3) is sharp, we cannot get a better estimate without some additional restrictions. (See Example 3.3 in [2]). The estimate (1.2) in the extension of Theorem 1 in the case $\bar{u}(\sigma) \leq v \leq \sqrt{n}\sigma^2$ is also sharp (we disregard again the value of the universal constants in this formula), since the tail-distribution of a normalized partial sum cannot have a better bound, than the Gaussian estimate given in (1.2). Formally there is a gap between the results of Theorem 1 and its extension, because we did not consider the case $\sqrt{n}\sigma^2 \leq v \leq u(\sigma)$. But this gap can be simply filled in the case when the numbers D and L are bounded by constants not depending on σ^2 , and we do not try to find optimal universal constants in our estimates. Indeed, in this case we have $u(\sigma) \leq \frac{\bar{C}}{\sqrt{n}}n\sigma^2$, and

$$P\left(\sup_{f \in \mathcal{F}} |S_n(f)| \geq v\right) \leq P\left(\sup_{f \in \mathcal{F}} |S_n(f)| \geq \sqrt{n}\sigma^2\right) \leq Ce^{-\alpha n\sigma^2} \leq Ce^{-\bar{\alpha}v^2/\sigma^2},$$

i.e. relation (1.2) holds (with a possible different parameter $\bar{\alpha} > 0$) for all $\bar{u}(\sigma) < v \leq u(\sigma)$. This estimate is sharp again.

We also have to understand why we could give a good estimate for the supremum of normalized partial sums only under the conditions $v \geq u(\sigma)$ in cases (a) and (b) and $v \geq \bar{u}(\sigma)$ in case (c). I shall present an example that satisfies the conditions of Theorem 1, and in which there is no useful estimate in formulas (1.1) and (1.2) for $v < u(\sigma)$ in cases (a) and (b) or $v < \bar{u}(\sigma)$ in cases (c). (More precisely, we allow a different multiplying factor C_j as in the definition of $u(\sigma)$ and $\bar{u}(\sigma)$ when we consider this model.) This implies in particular that the conditions $v \geq u(\sigma)$ and $v \geq \bar{u}(\sigma)$ cannot be dropped in Theorem 1 and in its extension.

At this point it may be useful to recall the concentration inequality for the supremum of partial sums of independent random variables. (See e.g. [5]). It states that there is a concentration point of the supremum of partial sums of independent random variable such that this supremum is strongly concentrated in a small neighbourhood of this concentration point. I do not formulate this result in a more precise and detailed form, because we need it here only for the sake of some orientation. The problem with the application of this result is that it determines the concentration point only in an implicit way as the expected value of the supremum we are investigating, and we cannot calculate it explicitly in the general case. On the other hand, the concentration inequality implies that we can get a good, non-trivial estimate for the tail distribution of partial sums of independent random variables only at levels higher than the concentration point of the partial sums. (We call such estimates trivial which only say that a probability is not greater than 1.) So the numbers $u(\sigma)$ and $\bar{u}(\sigma)$ in Theorem 1 and in its extension are actually upper bounds for the concentration point of the supremum, and we shall present a model satisfying the conditions of Theorem 1, where the values $u(\sigma)$ and $\bar{u}(\sigma)$ determine the concentration point of the supremum up to a multiplicative factor.

We shall consider the following model. Take independent, uniformly distributed random variables ξ_1, \dots, ξ_n on the unit interval $[0, 1]$, fix a number $0 \leq \sigma^2 \leq 1$, and define a class of functions \mathcal{F}_σ and $\bar{\mathcal{F}}_\sigma$ with functions defined on the unit interval $[0, 1]$ in the following way. $\mathcal{F}_\sigma = \{f_1, \dots, f_k\}$, and $\bar{\mathcal{F}} = \{\bar{f}_1, \dots, \bar{f}_k\}$ with $k = k(\sigma) = \lfloor \frac{1}{\sigma^2} \rfloor$, where $\lfloor \cdot \rfloor$ denotes integer part, and $\bar{f}_j(x) = f_j(x|\sigma) = 1$ if $x \in [(j-1)\sigma^2, j\sigma^2)$, $\bar{f}_j(x) = f_j(x|\sigma) = 0$ if $x \notin [(j-1)\sigma^2, j\sigma^2)$, $1 \leq j \leq k$, and $f_j(x) = f_j(x|\sigma) = \bar{f}_j(x) - \sigma^2$, $1 \leq j \leq n$. It can be seen that \mathcal{F}_σ satisfies the conditions of Theorem 1 with parameter σ^2 . In particular, it is an L_1 -dense class with such a parameter D and exponent L that can be bounded by numbers not depending on σ^2 . This can be seen directly, but it is also a consequence of some classical results by which the indicator functions of a Vapnik–Červonenkis class of sets constitute an L_1 dense class of functions. (See e.g. Theorem 5.2 in [2]).

I shall show that in this example a number $\bar{C} > 0$ can be chosen in such a way that for all $\delta > 0$ there is an index $n_0(\delta)$ such that for all sample sizes $n \geq n_0(\delta)$ and numbers $0 \leq \sigma \leq 1$ the inequality

$$P \left(\sup_{f \in \mathcal{F}_\sigma} |S_n(f)| \geq \hat{u}(\sigma) \right) \geq 1 - \delta, \quad (2.1)$$

holds with $\hat{u}(\sigma) = \frac{\bar{C}}{\sqrt{n}}$ in case (a), $\hat{u}(\sigma) = \frac{\bar{C}}{\sqrt{n}} \frac{\log n}{\log(\frac{\log n}{n\sigma^2})}$ in case (b), and $\hat{u}(\sigma) = \bar{C}\sigma \log^{1/2} \frac{2}{\sigma}$ in case (c). This result may explain why we had to impose the conditions $v > u(\sigma)$ and $v > \bar{u}(\sigma)$ in Theorem 1 and in its extension. (We are interested only in such cases when the estimate of Theorem 1 or its extension provide an upper bound strictly less than 1, (i.e. smaller than a number $\alpha < 1$ for all parameters σ^2 and n), and this is the case if the constants C_j , $j = 3, 4, 5, 6$ are chosen sufficiently large in these results.)

To prove relation (2.1) introduce the following notation. Define the empirical dis-

tribution function $F_n(x)$ of the random variables ξ_1, \dots, ξ_n , i.e. put

$$F_n(x) = \frac{1}{n} \{\text{the number of indices } j, 1 \leq j \leq n, \text{ such that } \xi_j < x\}$$

for all $0 < x \leq 1$, and take its normalization $G_n(x) = \sqrt{n}(F_n(x) - x)$, $0 < x \leq 1$. Observe that

$$\left\{ \sup_{f \in \mathcal{F}_\sigma} |S_n(f)| \geq \hat{u}(\sigma) \right\} = \left\{ \max_{1 \leq j \leq k(\sigma)} |G_n(j\sigma) - G_n((j-1)\sigma)| \geq \hat{u}(\sigma) \right\}. \quad (2.2)$$

By a classical results of probability theory, the normalized empirical distribution functions weakly converge to the Brownian bridge as $n \rightarrow \infty$. In our next considerations it will be also interesting that the modulus of continuity of a Brownian bridge, (which actually agrees with the modulus of continuity of a Wiener process) can be also calculated. (see e.g. [4]). By a similar, but simpler calculation we can estimate the probability of the event we get by replacing the normalized empirical distribution function $G_n(\cdot)$ by a Brownian bridge in the right-hand side expression of (2.2). This is actually done with the choice $u(\sigma) = \bar{C}\sigma \log^{1/2} \frac{2}{\sigma}$ in the fourth chapter of [2] (page 27), and it is shown that this probability is almost one for large parameters n for all $\sigma > 0$ if the coefficient \bar{C} of $u(\sigma)$ is chosen sufficiently small. (Actually we have to choose $\bar{C} < \sqrt{2}$.) Let us call this estimate the Gaussian version of formula (2.1). At a heuristic level this result together with formula (2.2) and the weak convergence of the normalized empirical processes $G_n(\cdot)$ to a Brownian bridge suggests that formula (2.1) should hold with $u(\sigma) = \bar{C}\sigma \log^{1/2} \frac{2}{\sigma}$ and a small coefficient $\bar{C} > 0$.

This heuristic argument is nevertheless misleading, since the weak convergence of the empirical processes $G_n(\cdot)$ to the Brownian bridge does not make possible to carry out a limiting procedure that leads to formula (2.1). On the other hand, a stronger version of the weak convergence of the normalized empirical processes (see [1]) yields a useful result in this direction. This result states a normalized empirical process $G_n(x)$ and a Brownian bridge $B(x)$, $0 \leq x \leq 1$, can be constructed in such a way that $\sup_{0 \leq x \leq 1} |B(x) - G_n(x)| \leq K \frac{\log n}{\sqrt{n}}$ for all $n \geq 2$ and sufficiently large $K > 0$ with probability almost 1. This result together with the Gaussian version of formula (2.1) imply the validity of formula (2.1) if $\sigma^2 \geq B \frac{\log n}{2n}$ with a sufficiently large $B > 0$. Indeed, in this case $\hat{u}(\sigma) \geq 2K \frac{\log n}{\sqrt{n}}$, hence the Gaussian version of formula of (2.1) together with the result of [1] imply that

$$P \left(\max_{1 \leq j \leq k(\sigma)} |G_n(j\sigma) - G_n((j-1)\sigma)| \geq \frac{\hat{u}(\sigma)}{2} \right) \geq 1 - \delta$$

if $\sigma^2 \geq B \frac{\log n}{n}$, and $n \geq n_0(\delta)$, i.e. inequality (2.1) holds in this case if we replace \bar{C} by $\frac{\bar{C}}{2}$ in the definition of $\hat{u}(\sigma)$. Moreover, this relation holds for all $\sigma^2 \geq \frac{\log n}{8n}$, i.e. in the case (c) generally if we choose $\hat{u}(\sigma) = \bar{C}\sigma \log^{1/2} \frac{2}{\sigma}$ with a sufficiently small $\bar{C} > 0$. To

see this it is enough to observe that if $\max_{1 \leq j \leq k(\sigma)} |G_n(j\sigma) - G_n((j-1)\sigma)| \leq \hat{u}(\sigma)$, then for any positive integers A we have $\max_{1 \leq j \leq k(\sqrt{A}\sigma)} |G_n(j(A\sigma)) - G_n((j-1)(A\sigma))| \leq A\hat{u}(\sigma)$, and that the corresponding result holds if $\sigma^2 \geq B \frac{\log n}{8n}$.

In cases (a) and (b) the above Gaussian approximation argument does not work. In case (b) we shall prove formula (2.1) by means of a Poissonian approximation method described below. It can be considered as a more detailed elaboration of the argument in Example 4.3 of [2].

In this argument first we consider the following problem. Take a Poisson process $Z_n(t)$, $0 \leq t \leq 1$, with parameter n , (i.e. let $E Z_n(t) = nt$ for all $0 \leq t \leq 1$) in the interval $[0, 1]$. Fix some number $0 \leq \sigma^2 \leq \frac{1}{7} \frac{\log n}{n}$, and define with its help the number $\hat{u}(\sigma) = \hat{u}(\sigma, n) = \frac{3}{4\sqrt{n}} \frac{\log n}{\log(\frac{\log n}{n\sigma^2})}$ and the random variables $\bar{V}_j = \bar{V}_j^{(n)}(\sigma) = Z_n(j\sigma^2) - Z_n((j-1)\sigma^2)$ for $1 \leq j \leq k$ with $k = \lfloor \frac{1}{\sigma^2} \rfloor$. (Here we defined $\hat{u}(\sigma)$ similarly to quantity introduced with the same notation at the formulation of inequality (2.1) in the case (b). We only made small modifications. Namely we considered σ^2 in the interval $[0, \frac{1}{7} \frac{\log n}{n}]$ instead of the interval $[\frac{1}{n^{200}}, \frac{\log n}{8n}]$, and we fixed the value $\bar{C} = \frac{3}{4}$ in the definition of $\hat{u}(\sigma)$. We want to show that for all $\delta > 0$ there is some threshold index $n_0(\delta)$ such that the inequality

$$P\left(\max_{1 \leq j \leq k(\sigma)} \bar{V}_j^{(n)}(\sigma) \geq \sqrt{n}\hat{u}(\sigma, n)\right) \geq 1 - \delta \quad \text{if } n \geq n_0(\delta) \quad (2.3)$$

holds for all $0 \leq \sigma^2 \leq \frac{1}{7} \frac{\log n}{n}$.

To prove this inequality let us first observe that

$$\begin{aligned} P\left(\max_{1 \leq j \leq k(\sigma)} \bar{V}_j^{(n)}(\sigma) \geq \sqrt{n}\hat{u}(\sigma, n)\right) &\geq P(\bar{V}_j^{(n)}(\sigma) = \sqrt{n}\hat{u}(\sigma, n) \text{ for some } 1 \leq j \leq k) \\ &= 1 - P(\bar{V}_1^{(n)}(\sigma) \neq \sqrt{n}\hat{u}(\sigma, n))^k, \end{aligned}$$

and

$$\begin{aligned} P(\bar{V}_1^{(n)}(\sigma) \neq \sqrt{n}\hat{u}(\sigma, n)) &= 1 - P(\bar{V}_1^{(n)}(\sigma) = \sqrt{n}\hat{u}(\sigma, n)) \\ &= 1 - \frac{(n\sigma^2)^{\sqrt{n}\hat{u}(\sigma, n)}}{(\sqrt{n}\hat{u}(\sigma, n))!} e^{-n\sigma^2} \leq 1 - \left(\frac{n\sigma^2}{\sqrt{n}\hat{u}(\sigma, n)}\right)^{\sqrt{n}\hat{u}(\sigma, n)} e^{-n\sigma^2}. \end{aligned}$$

Since we have $k = \frac{1}{\sigma^2}$ we can bound the left-hand side of (2.3) from below as

$$P\left(\max_{1 \leq j \leq k(\sigma)} \bar{V}_j^{(n)}(\sigma) \geq \sqrt{n}\hat{u}(\sigma, n)\right) \geq 1 - \left[1 - \left(\frac{n\sigma^2}{\sqrt{n}\hat{u}(\sigma, n)}\right)^{\sqrt{n}\hat{u}(\sigma, n)} e^{-n\sigma^2}\right]^{1/\sigma^2} \geq 1 - e^{-T}$$

with $T = \frac{1}{\sigma^2} \left(\frac{n\sigma^2}{\sqrt{n}\hat{u}(\sigma, n)}\right)^{\sqrt{n}\hat{u}(\sigma, n)} e^{-n\sigma^2}$, hence to prove (2.3) it is enough to show that

$$\left(\frac{n\sigma^2}{\sqrt{n}\hat{u}(\sigma, n)}\right)^{\sqrt{n}\hat{u}(\sigma, n)} \geq \sigma^2 e^{n\sigma^2} \log \frac{1}{\delta} \quad \text{if } n \geq n_0(\delta). \quad (2.4)$$

The right-hand side of (2.4) can be bounded from above as

$$\sigma^2 e^{n\sigma^2} \log \frac{1}{\delta} = \frac{\log \frac{1}{\delta}}{n} (n\sigma^2) e^{n\sigma^2} \leq \frac{\log \frac{1}{\delta}}{n} \left(\frac{1}{7} \log n \right) e^{(\log n)/7} \leq n^{-5/6}$$

if $n \geq n_0(\delta)$, since $n\sigma^2 \leq \frac{1}{7} \log n$. Hence we prove (2.4) if we show that

$$\frac{\sqrt{n}\hat{u}(n, \sigma)}{n\sigma^2} \log \left(\frac{\sqrt{n}\hat{u}(\sigma, n)}{n\sigma^2} \right) \leq \frac{5 \log n}{6 n\sigma^2}.$$

By applying the definition of $\hat{u}(n, \sigma)$ and introducing the quantity $z = \frac{3}{4} \frac{\log n}{n\sigma^2}$ we can rewrite the last inequality as $\frac{z}{\log(\frac{4z}{3})} \log(\frac{z}{\log(\frac{4z}{3})}) \leq \frac{10}{9}z$, or since $z \geq \frac{21}{4}$ in the case we are investigating it can be rewritten as $\frac{1}{9} \log \frac{4z}{3} \geq -\log \log \frac{4z}{3} - \log \frac{4}{3}$ if $z \geq \frac{21}{4}$, and this relation clearly holds. Thus we proved (2.3).

We shall prove relation (2.1) in the case (b) by means of formula (2.3) for a Poisson process with parameter $\frac{99}{100}n$ instead of n and a simple coupling argument between an empirical process and a Poisson process. Namely, we make the following coupling. Let us consider a sequence of independent random variables ξ_1, ξ_2, \dots with uniform distribution on the unit interval $[0, 1]$ together with a Poissonian random variable $\eta = \eta_n$ with parameter $\frac{99}{100}n$ independent of the random variables ξ_j , $j = 1, 2, \dots$, and take the first η_n terms of the random variables ξ_j , i.e. the sequence $\xi_1, \xi_2, \dots, \xi_{\eta_n}$ with the random stopping index η_n . In such a way we constructed a Poisson process with parameter $\frac{99}{100}n$, which is smaller than the (non-normalized) empirical distribution of the sequence ξ_1, \dots, ξ_n in the following sense. For large parameter n with probability almost 1 all intervals $[a, b] \subset [0, 1]$ contain more points from the sequence ξ_1, \dots, ξ_n than from the above constructed Poisson process. This is a simple consequence of the fact that $P(\eta_n > n) \rightarrow 0$ as $n \rightarrow \infty$.

The above coupling construction and formula (2.3) (with a Poisson process with parameter $\frac{99}{100}$) imply that

$$P \left(\sup_{\bar{f} \in \bar{\mathcal{F}}_\sigma} \sqrt{n} S_n(\bar{f}) \geq \sqrt{\frac{99}{100}} n \hat{u}(\sigma, \left(\frac{99}{100} n \right)) \right) \geq 1 - \delta \quad \text{if } n \geq n_0(\delta)$$

with the class of functions $\bar{\mathcal{F}}_\sigma$ introduced before the formulation (2.1) and the function $\hat{u}(\sigma, n)$ defined in the discussion of case (b). To complete the proof of (2.1) in the case (b) it is enough to check that the above relation remains valid if the class of functions $\bar{\mathcal{F}}_\sigma$ is replaced by the class of functions \mathcal{F}_σ and the term $\sqrt{\frac{99}{100}} n \hat{u}(\sigma, \frac{99}{100} n)$ is replaced by $\hat{u}(\sigma, n) = \frac{\bar{C}}{\sqrt{n}} \frac{\log n}{\log(\frac{\log n}{n\sigma^2})}$ with some appropriate $\bar{C} > 0$. Since the functions $f \in \mathcal{F}$ are of the form $f(x) = \bar{f}(x) - \sigma^2$ with some $\bar{f} \in \mathcal{F}$, this has the consequence $\sqrt{n} S_n(f) = \sqrt{n} S_n(\bar{f}) - n\sigma^2$, and to prove the desired relation it is enough to check that

$$\sqrt{\frac{99}{100}} \frac{3}{4} \frac{\log n}{\log(\frac{\log n}{\frac{99}{100} n\sigma^2})} - n\sigma^2 \geq \sqrt{\frac{99}{100}} \frac{3}{4} \frac{\log n}{\log(\frac{\log n}{n\sigma^2})} - n\sigma^2 \geq \bar{C} \frac{\log n}{\log(\frac{\log n}{n\sigma^2})}$$

with some appropriate $\bar{C} > 0$ if $8n\sigma^2 \leq \log n$. The first inequality clearly holds, and the second inequality is equivalent to the relation

$$\sqrt{\frac{99}{100}} \frac{3}{4} \frac{\frac{\log n}{n\sigma^2}}{\log(\frac{\log n}{n\sigma^2})} \geq \alpha$$

with some $\alpha > 1$. But this relation clearly holds if $8n\sigma^2 \leq \log n$. Thus we have proved (2.1) also in case (b).

In the case (a) the proof of (2.1) is very simple. It is enough to observe that the sample points ξ_j fall into one of the intervals $[(j-1)\sigma^2, j\sigma^2)$, $1 \leq j \leq k$, (we disregard the event that they fall into the last interval $[k\sigma^2, 1)$ which has negligible small probability), hence

$$P \left(\sup_{\bar{f} \in \bar{\mathcal{F}}_\sigma} \sqrt{n} S_n(\bar{f}) = 1 \right) \geq 1 - \delta \quad \text{if } n \geq n_0(\delta),$$

and since σ^2 is very small for large n relation (2.1) holds in case (a) with $\bar{C} = 1 - \varepsilon$ for any $\varepsilon > 0$.

At the end of this section let me remark that in the above example actually we have given a lower bound on the modulus of continuity of a normalized empirical process. I formulate a problem below where the proof of a stronger form of this result is suggested.

Problem. *Let ξ_1, ξ_2, \dots be a sequence of independent random variables, uniformly distributed in the unit interval $[0, 1]$, and define with its help the empirical distribution functions*

$$F_n(x) = \frac{1}{n} \text{ times the number of indices } j, 1 \leq j \leq n, \text{ such that } \xi_j < x$$

together with their normalizations $G_n(x) = \sqrt{n}(F_n(x) - x)$, $0 \leq x \leq 1$, for all indices $n = 1, 2, \dots$. Find such a function $v(n, \sigma^2)$, $n = 1, 2, \dots$, $0 \leq \sigma^2 \leq 1$, for which

$$\lim_{n \rightarrow \infty} \sup_{\{(s,t): 0 \leq s, t \leq 1, |t-s| \leq \sigma_n^2\}} \frac{|G_n(t) - G_n(s)|}{v(n, \sigma_n^2)} = 1 \quad \text{with probability 1}$$

if $\sigma_n^2 \rightarrow 0$ as $n \rightarrow \infty$.

3. Proof of Theorem 1 and its extension.

Proof of Theorem 1. In the case (a) inequality (1.1) is a simple consequence of Theorem 1 in [3]. We can apply this result (by writing σ^2 instead of ρ in its formulation), since $\int f^2(x)\mu(dx) \leq \int |f(x)|\mu(dx)$ if $\sup_{x \in X} |f(x)| \leq 1$, hence under the conditions of Theorem 1 the inequality $\int |f(x)|\mu(dx) \leq \rho$ holds for all $f \in \mathcal{F}$ with $\rho = \sigma^2$. Hence

$$P\left(\sup_{f \in \mathcal{F}} |S_n(f)| \geq v\right) \leq D e^{-\frac{1}{50}\sqrt{nv} \log(\sigma^{-2})} \quad \text{if } v \geq \frac{\bar{C}}{\sqrt{n}}L \text{ and } \sigma^2 \leq \frac{1}{n^{200}} \quad (3.1)$$

with an appropriate $\bar{C} > 0$.

I claim that we can drop the coefficient D at the right-hand side of (3.1) if we replace the coefficient $\frac{1}{50}$ by $\frac{1}{100}$ in the exponent, we choose such a constant \bar{C} in (3.1) for which $\bar{C} \geq \frac{1}{2}$, and impose condition (a) in the form $v \geq \frac{\bar{C}}{\sqrt{n}}(L + \frac{\log D}{\log n})$. To show this it is enough to check that $D \leq e^{\frac{1}{100}\sqrt{nv} \log(\sigma^{-2})}$ in this case. This relation holds, since $\frac{\log D}{\log n} \leq 2\sqrt{nv}$, and $\log(\sigma^{-2}) \geq 200 \log n$, thus $D = \exp\{\frac{1}{200}(\frac{\log D}{\log n})(200 \log n)\} \leq \exp\{\frac{1}{100}\sqrt{nv} \log(\sigma^{-2})\}$, as I claimed.

Next I show that formula (3.1) or its previous modification remains valid if we replace $\log(\sigma^{-2})$ by $\log(\frac{v}{\sqrt{n}\sigma^2})$ in the exponent of its right-hand side. In the proof of this statement we can restrict our attention to the case $v \leq \sqrt{n}$, since otherwise the probability at the left-hand side of (3.1) equals zero. In this case the inequality $\sigma^{-2} \geq \frac{v}{\sqrt{n}\sigma^2}$ holds, and this allows the above replacement. The above modifications of formula (3.1) imply inequality (1.1) in case (a).

Remark. If we are not interested in the value of the (universal) constants in (1.1), then in the case (a) this inequality has the same strength if we replace the term $\log(v/\sqrt{n}\sigma^2)$ by $\log(\sigma^{-2})$ in it. To see this, observe that beside the inequality $\sigma^{-2} \geq \frac{v}{\sqrt{n}\sigma^2}$ (if $v \leq \sqrt{n}$), the inequality $\frac{v}{\sqrt{n}\sigma^2} \geq \frac{1}{n\sigma^2} \geq \sigma^{-2+1/100}$ also holds in case (a) because of the inequalities $v \geq u(\sigma) \geq n^{-1/2}$ and $n^{-200} \geq \sigma^2$. The original form of (1.1) has the advantage that it simultaneously holds in all cases (a), (b) and (c).

The proof of Theorem 1 in cases (b) and (c). By applying the L_1 -dense property of the class of functions \mathcal{F} with the choice $\varepsilon = n^{-200}$ and the measure μ we may find $m \leq Dn^{200L}$ functions $f_j \in \mathcal{F}$, $1 \leq j \leq m$, such that $\min_{1 \leq j \leq m} \int |f_j(x) - f(x)|\mu(dx) \leq n^{-1/200}$

for all $f \in \mathcal{F}$. This means that $\mathcal{F} = \bigcup_{j=1}^n \mathcal{D}_j$ with

$$\mathcal{D}_j = \left\{ f: f \in \mathcal{F}, \int |f_j(x) - f(x)|\mu(dx) \leq n^{-200} \right\},$$

and as a consequence

$$P\left(\sup_{f \in \mathcal{F}} |S_n(f)| \geq v\right) \leq \sum_{j=1}^m P\left(|S_n(f_j)| \geq \frac{v}{2}\right) + \sum_{j=1}^m P\left(\sup_{f \in \mathcal{D}_j} |S_n(f - f_j)| \geq \frac{v}{2}\right) \quad (3.2)$$

for all $v > 0$. We shall estimate both terms at the right-hand side of (3.2) if $v \geq u(\sigma)$, the first one by means of Bennett's inequality, more precisely by a consequence of this inequality formulated below, and the second term by means of the already proved case (a) of Theorem 1. We shall apply the following version of Bennett's inequality, see [2].

Bennett's inequality. *Let X_1, \dots, X_n be independent and identically distributed random variables such that, $P(|X_1| \leq 1) = 1$, $EX_1 = 0$, and $EX_1^2 \leq \sigma^2$ with some $0 \leq \sigma \leq 1$. Put $S_n = \frac{1}{\sqrt{n}} \sum_{j=1}^n X_j$. Then*

$$P(S_n > v) \leq \exp \left\{ -n\sigma^2 \left[\left(1 + \frac{v}{\sqrt{n}\sigma^2} \right) \log \left(1 + \frac{v}{\sqrt{n}\sigma^2} \right) - \frac{v}{\sqrt{n}\sigma^2} \right] \right\} \quad \text{for all } v > 0.$$

As a consequence, for all $\varepsilon > 0$ there exists some $B = B(\varepsilon) > 0$ such that

$$P(S_n > v) \leq \exp \left\{ -(1 - \varepsilon)\sqrt{n}v \log \frac{v}{\sqrt{n}\sigma^2} \right\} \quad \text{if } v > B\sqrt{n}\sigma^2,$$

and there exists some positive constant $K > 0$ such that

$$P(S_n > v) \leq \exp \left\{ -K\sqrt{n}v \log \frac{v}{\sqrt{n}\sigma^2} \right\} \quad \text{if } v > 2\sqrt{n}\sigma^2. \quad (3.3)$$

The above result is a special case of Theorem 3.2 in [2], in the case when we restrict our attention to sums of independent and identically distributed random variables. It has a slightly different form, because in the definition of S_n we considered normalized sums (with a multiplication by $n^{-1/2}$). Here we need only the inequality formulated in (3.3) which helps to estimate the probabilities appearing in the first sum at the right-hand side of (3.2). To apply (3.3) in the estimation of these terms we have to show that $u(\sigma) > 2\sqrt{n}\sigma^2$ in cases (b) and (c) if the constants C_4 and C_5 are chosen sufficiently large in Theorem 1.

In case (b) it is enough to show that $u(\sigma) \geq C_4 \frac{\log n}{\log(\frac{\log n}{n\sigma^2})} \geq 2n\sigma^2$, and even $C_4 \frac{\log n}{\log(\frac{\log n}{n\sigma^2})} \geq 20n\sigma^2$, or in an equivalent form $\frac{C_4}{20} \frac{\log n}{n\sigma^2} \geq \log(\frac{\log n}{n\sigma^2})$. (Observe that $\frac{\log n}{n\sigma^2} \geq 8$, hence $\log(\frac{\log n}{n\sigma^2}) > 0$ in case (b).) This statement holds, since $z = \frac{\log n}{n\sigma^2} \geq 2$ in case (b), and $\frac{C_4}{20} z \geq \log z$ if $z \geq 8$, and C_4 is sufficiently large.

In case (c), clearly $u(\sigma) \geq \frac{C_5}{\sqrt{n}} n\sigma^2 \geq 20\sqrt{n}\sigma^2$ for sufficiently large constant C_5 . These relations together with formula (3.3) imply that in cases (b) and (c)

$$P\left(|S_n(f_j)| \geq \frac{v}{2}\right) \leq 2 \exp \left\{ -K\sqrt{n}v \log \frac{v}{\sqrt{n}\sigma^2} \right\} \quad \text{if } v \geq u(\sigma) \quad (3.4)$$

with an appropriate $K > 0$ for all $1 \leq j \leq m$. (In formula (3.4) we exploit that $\log(\frac{v}{\sqrt{n}\sigma^2}) \geq \frac{1}{2} \log(\frac{v}{\sqrt{n}\sigma^2})$ since $\frac{v}{\sqrt{n}\sigma^2} \geq 20$, and as a consequence $\log(\frac{v}{\sqrt{n}\sigma^2}) \geq 2 \log 2$.)

Let us define, with the help of the class of functions \mathcal{D}_j the class of functions $\mathcal{D}'_j = \{h: h = \frac{f-f_j}{2}, f \in \mathcal{D}_j\}$ for all $1 \leq j \leq m$. It is not difficult to see that $\sup_{x \in X} |h(x)| \leq 1$, $\int h^2(x) \mu(dx) \leq \int |h(x)| \mu(dx) \leq n^{-200}$ for all $h \in \mathcal{D}'_j$, and \mathcal{D}'_j is an L_1 -dense class of functions with parameter D and exponent L , $1 \leq j \leq m$. I claim that

$$\begin{aligned} P \left(\sup_{f \in \mathcal{D}_j} |S_n(f - f_j)| \geq \frac{v}{2} \right) &= P \left(\sup_{h \in \mathcal{D}'_j} |S_n(h_j)| \geq \frac{v}{4} \right) \\ &\leq e^{-C_2 \sqrt{n} v \log(v n^{195})} \quad \text{if } v \geq u(\sigma) \end{aligned} \quad (3.5)$$

for all $1 \leq j \leq n$ in both cases (b) and (c). We shall get this estimate by applying Theorem 1 in the already proved case (a) with the choice of parameter $\sigma^2 = n^{-200}$. To apply this result we have to check that $\frac{u(\sigma)}{4} \geq u(n^{-200}) = \frac{C_3}{\sqrt{n}}(L + \frac{\log D}{\log n})$ if the constants C_4 and C_5 are sufficiently large. These statements hold, since in case (b) $\frac{\log n}{\log \frac{\log n}{n \sigma^2}} \geq \frac{\log n}{\log(n^{199} \log n)} \geq \frac{1}{200}$, hence $u(\sigma) \geq \frac{C_4}{\sqrt{n}}(\frac{L \log n}{200} + \log D) \geq 4u(n^{-200})$ if C_4 is chosen sufficiently large, and an analogous but simpler argument supplies this relation in case (c) if C_5 is chosen sufficiently large.

It is not difficult to see that the right-hand side both of (3.4) and (3.5) can be bounded from above by $C_1 e^{-\bar{C}_2 \sqrt{n} v \log(v/\sqrt{n} \sigma^2)}$ with some appropriate constants $C_1 > 0$ and $\bar{C}_2 > 0$. Hence relations (3.2), (3.4) and (3.5) together with the inequality $m \leq D n^{200L}$ imply that

$$P \left(\sup_{f \in \mathcal{F}} |S_n(f)| \geq v \right) \leq C_1 D n^{200L} e^{-\bar{C}_2 \sqrt{n} v \log(v/\sqrt{n} \sigma^2)} \quad \text{if } v \geq u(\sigma) \quad (3.6)$$

in both cases (b) and (c). Hence to complete the proof of Theorem 1 (with the choice $C_2 = \frac{\bar{C}_2}{2}$) it is enough to show that

$$e^{-\frac{\bar{C}_2}{2} \sqrt{n} v(\sigma) v / \sqrt{n} \sigma^2} \leq e^{-\frac{\bar{C}_2}{2} \sqrt{n} u(\sigma) \log(u(\sigma) / \sqrt{n} \sigma^2)} \leq D^{-1} n^{-200L} \quad \text{if } v \geq u(\sigma) \quad (3.7)$$

in cases (b) and (c) if the constants C_4 and C_5 are chosen sufficiently large.

It is enough to prove the second inequality in formula (3.7), since its proof also implies that the expressions in the exponent of this formula have negative value, and they are decreasing functions for $v \geq u(\sigma)$. The second inequality in (3.7) clearly holds in case (c), since $\frac{\bar{C}_2}{2} \sqrt{n} u(\sigma) \geq 200L \log n + \log D$, and $\frac{\log(u(\sigma))}{\sqrt{n} \sigma^2} \geq 1$ in this case. In case (b) relation (3.7) can be reduced to the inequalities $\bar{C}_2 \sqrt{n} u(\sigma) \log(\frac{u(\sigma)}{\sqrt{n} \sigma^2}) \geq 800L \log n$, and $\bar{C}_2 \sqrt{n} u(\sigma) \log(\frac{\sqrt{n} u(\sigma)}{n \sigma^2}) \geq 4 \log D$. To prove the second inequality observe that in case (b)

$$\bar{C}_2 \sqrt{n} u(\sigma) \geq C_4 \bar{C}_2 \log D \geq 4 \log D, \quad \text{and} \quad \log \left(\frac{\sqrt{n} u(\sigma)}{n \sigma^2} \right) \geq 1.$$

The second of these inequalities follows from the relation $\frac{\sqrt{nu}(\sigma)}{n\sigma^2} \geq C_4 \frac{\frac{\log n}{n\sigma^2}}{\log(\frac{\log n}{n\sigma^2})} \geq 3$, which holds because of the relation $\frac{\log n}{n\sigma^2} \geq 8$ in case (b).

The remaining inequality can be rewritten as $\bar{C}_2 \frac{\sqrt{nu}(\sigma)}{n\sigma^2} \log(\frac{\sqrt{nu}(\sigma)}{n\sigma^2}) \geq 800L \frac{\log n}{n\sigma^2}$. To prove it observe that because of the definition of the function $u(\sigma)$ in case (b) we can write $\bar{C}_2 \frac{\sqrt{nu}(\sigma)}{n\sigma^2} \geq 1600L \frac{\log n}{n\sigma^2} \frac{1}{\log(\frac{\log n}{n\sigma^2})} \geq 1600L \frac{\log n}{n\sigma^2}$, since $\log(\frac{\log n}{n\sigma^2}) \geq \log 8 \geq 1$. I also claim that $\log(\frac{\sqrt{nu}(\sigma)}{n\sigma^2}) \geq \frac{1}{2}(\frac{\log n}{n\sigma^2})$. By multiplying the last two inequalities we get the desired inequality, and this completes the proof of Theorem 1.

To prove the above formulated inequality introduce the notation $z = \frac{\log n}{n\sigma^2}$. By exploiting the definition of $u(\sigma)$ in case (b) we can write with the help of this notation that $\log(\frac{\sqrt{nu}(\sigma)}{n\sigma^2}) \geq \log z - \log \log z \geq \frac{1}{2} \log z = \frac{1}{2}(\frac{\log n}{n\sigma^2})$. In the above argument we have exploited that in case (b) $z \geq 8$, hence $\log z - \log \log z \geq \frac{1}{2} \log z$. Theorem 1 is proved.

The extension of Theorem 1 is actually a reformulation of Theorem 4.1 in [2], and its proof is worked out there in detail. Nevertheless, I briefly discuss this result to get a better understanding of it. Its proof is based on two propositions, and one of them is actually a weakened version of Theorem 1 of this paper.

On the proof of the extension of Theorem 1. This result is proved in [2] with the help of two results formulated in Propositions 6.1 and 6.2 of that work. I discuss their content, and show that Proposition 6.2 is a weakened version of Theorem 1. First I reformulate a slightly modified version of it in the following Theorem 3.1.

Theorem 3.1. *Let us have a probability measure μ on a measurable space (X, \mathcal{X}) together with a sequence of independent and μ distributed random variables ξ_1, \dots, ξ_n , $n \geq 2$, and a countable, L_1 -dense class \mathcal{F} of functions $f = f(x)$ on (X, \mathcal{X}) with some parameter $D \geq 1$ and exponent $L \geq 1$ which satisfies the conditions $\sup_{x \in X} |f(x)| \leq 1$, $\int f(x)\mu(dx) = 0$ and $\int f^2(x)\mu(dx) \leq \sigma^2$ for all $f \in \mathcal{F}$ with some $0 < \sigma \leq 1$ such that the inequality $n\sigma^2 > L \log n + \log D$ holds. Then there exists a threshold index A_0 such that the normalized random sums $S_n(f)$, $f \in \mathcal{F}$, introduced in Theorem 1 satisfy the inequality*

$$P \left(\sup_{f \in \mathcal{F}} |S_n(f)| \geq An^{1/2}\sigma^2 \right) \leq e^{-A^{1/2}n\sigma^2/2} \quad \text{if } A \geq A_0. \quad (3.8)$$

I show that the estimate (3.8) in Theorem 3.1 is a weakened version of formula (1.1) of Theorem 1. First I show that the probability at the left-hand side of (3.8) can be estimated by means of Theorem 1 in case (c) with the choice $v = An^{1/2}\sigma^2$ if $A \geq A_0$ with a sufficiently large threshold index $A_0 > 0$. We have to check that $v \geq u(\sigma)$ if A_0 is chosen sufficiently large. But under the conditions of Theorem 3.1 $n\sigma^2 \geq L \log n \geq \frac{1}{8}n\sigma^2$, and for $v = An^{1/2}\sigma^2$ we can write $v \geq \frac{A_0}{\sqrt{n}}n\sigma^2 \geq \frac{A_0}{2\sqrt{n}}n\sigma^2 + \frac{A_0}{2\sqrt{n}}(L \log n + \log D) \geq \frac{C_5}{\sqrt{n}}(n\sigma^2 + L \log n + \log D) = u(\sigma)$.

Thus we can apply formula (1.1) with $v = An^{1/2}\sigma^2$ to estimate the left-hand side of (3.8), and we get the upper bound $C_1 e^{-C_2 \sqrt{nv} \log(v/\sqrt{n}\sigma^2)} = C_1 e^{-C_2 An\sigma^2 \log A}$ if $A \geq A_0$. This is an estimate sharper than formula (3.8) if the (universal) constant A_0 is chosen sufficiently large. This calculation also indicates that Theorem 3.1 provides such a good estimate as Theorem 1 if $A \leq \bar{A}_0$ with a fixed universal constant \bar{A}_0 . (Here we are not interested in the value of the universal constants in our estimates.) Besides, to prove the extension of Theorem 1 it is enough to have good estimates only in this case.

I shall only briefly discuss the content of Proposition 6.1 in [2], the other main ingredient in the proof of the extension of Theorem 1. Its proof is based on a classical method, called the chaining argument in the literature. It provides a sharp estimate for the tail distribution of the supremum of Gaussian random variables. But if we are interested in the tail distribution of the supremum of normalized partial sums of independent and identically distributed random variables, like in the extension of Theorem 1, then it only provides a weaker estimate. Proposition 6.1 actually contains the result we can get in our case with the help of the chaining method. This result is not sufficient for our purposes, but its combination with Theorem 3.1 enables us to prove the extension of Theorem 1.

Here I do not discuss the details of the chaining argument. It has a fairly detailed description in [6], but also [2] may help in understanding this method. I only remark that this method supplies a weaker estimate for the supremum of normalized partial sums of i.i.d. random variables, than for the supremum of the Gaussian random variables, because the tail distribution of partial sums of independent random variables has a slightly worse behaviour than the Gaussian tail distribution.

The main result of Proposition 6.1 in [2] states that under the conditions of the extension of Theorem 1 there exists such a set of functions $\mathcal{F}_{\bar{\sigma}} \subset \mathcal{F}$ with some nice properties for which the inequality

$$P\left(\sup_{f \in \mathcal{F}_{\bar{\sigma}}} |S_n(f)| \geq \frac{u}{\bar{A}}\right) \leq 4 \exp\left\{-\alpha \left(\frac{u}{10\bar{A}\sigma}\right)^2\right\} \quad (3.9)$$

holds if the number u satisfies the condition $n\sigma^2 \geq (\frac{u}{\sigma})^2 \geq C_6(L \log \frac{2}{\bar{\sigma}} + \log D)$ with a fixed constant $\bar{A} \geq 1$, and the (sufficiently large) number $C_6 = C_6(\bar{A})$ appearing in the condition of formula (3.9) depends on it. (The number \bar{A} was introduced in this estimate because of some technical reasons.) Moreover, the class of functions $\mathcal{F}_{\bar{\sigma}}$, where $\bar{\sigma} = \bar{\sigma}(u)$ depends on the number u in the above estimate has some properties which can be interpreted so that $\mathcal{F}_{\bar{\sigma}}$ is a relatively small and dense subset of \mathcal{F} . The set $\mathcal{F}_{\bar{\sigma}} = \{f_1, \dots, f_m\}$ has $m \leq D\bar{\sigma}^{-L}$ elements, and the sets $\mathcal{D}_j = \{f: f \in \mathcal{F}, \int (f - f_j)^2 d\mu \leq \bar{\sigma}^2\}$, $1 \leq j \leq m$, cover the set \mathcal{F} , i.e. $\bigcup_{j=1}^m \mathcal{D}_j = \mathcal{F}$. Theorem 6.1 also provides some control on $\bar{\sigma}$. Namely, $\frac{1}{16}(\frac{u}{\bar{A}\bar{\sigma}})^2 \geq n\bar{\sigma}^2 \geq \frac{1}{64}(\frac{u}{\bar{A}\bar{\sigma}})^2$, and the inequality $n\bar{\sigma}^2 \geq L \log n + \log D$ also holds.

Formula (3.9) gives a good (Gaussian type) estimate for a supremum of partial sums. But in this estimate we took the supremum for a class of functions $\mathcal{F}_{\bar{\sigma}} \subset \mathcal{F}$

instead of the class of functions \mathcal{F} . The chaining argument does not enable us to give a good estimate if we take the supremum for a subclass of \mathcal{F} much larger than $\mathcal{F}_{\bar{\sigma}}$. On the other hand, we get good estimates leading to the proof of the extension of Theorem 1 with the help of Proposition 6.1 and a good bound on the probabilities

$P\left(\sup_{h \in \mathcal{D}'_j} S_n(h) \geq \frac{u}{2}\right)$, $1 \leq j \leq m$, where $\mathcal{D}'_j = \{h = f - f_j: f_j \in \mathcal{D}_j\}$. Such a good

bound can be obtained with the help of Theorem 3.1. But to apply this result we have to know that $n\bar{\sigma}^2 \geq L \log n + \log D$, and this is the reason why this relation had to be proved in Proposition 6.1 of [2]. To get the desired estimate we also have to show that the number m of the sets \mathcal{D}_j is not too large. Since $m \leq D\bar{\sigma}^{-L}$ this can be proved with the help of the additional estimates on $\bar{\sigma}$ in Proposition 6.1. The proof of the extension of Theorem 1 with the help of Propositions 6.1 and 6.2 of [2] is contained in that work, so I omit the details. They show some similarity to the final step of the proof of Theorem 1 in this paper.

I finish this paper with the formulation of some comments and problems.

4. Some comments on the methods and results of this paper.

Our goal in this paper was to give a sharp estimate for the supremum of normalized partial sums $S_n(f)$, $f \in \mathcal{F}$, of i.i.d. random variables in such cases that were not covered by previous results. The classical methods, like the chaining argument do not work in the study of such problems, since we have to bound some events in their applications for which we cannot give a sufficiently good estimate. We have to deal with events whose probabilities are much larger than the value suggested by a Gaussian comparison. I wanted to find a method that would work also in such cases.

One natural candidate for it was the so-called symmetrization argument. The extension of Theorem 1 presented in this paper was proved with the help of this method in [2]. In the application of this method we consider a sequence of independent random variables $\varepsilon_1, \dots, \varepsilon_n$ with binomial distribution, i.e. $P(\varepsilon_j = 1) = P(\varepsilon_j = -1) = \frac{1}{2}$, $1 \leq j \leq n$, which is independent also of the random variables ξ_1, \dots, ξ_n , and we reduce, with

the help of some non-trivial inequalities, the estimation of $P\left(\sup_{f \in \mathcal{F}} \frac{1}{\sqrt{n}} \sum_{j=1}^n f(\xi_j) > v\right)$

to the estimation of its ‘symmetrized version’ $P\left(\sup_{f \in \mathcal{F}} \frac{1}{\sqrt{n}} \sum_{j=1}^n \varepsilon_j f(\xi_j) > v\right)$.

My original plan was to prove Theorem 1 by means of a refinement of the symmetrization argument. But I met hard problems when I tried to carry out this program. The proof with the help of the symmetrization argument would have required the application of such an induction procedure, where at the start we need a good estimate for the supremum of the normalized partial sums $S_n(f)$ of very small random terms $f(\xi_j)$. More explicitly we should have handled the case when the expectation of the absolute value $E|f(\xi_j)|$ of the terms in the sum are very small for all $f \in \mathcal{F}$. In such cases the symmetrization argument is not useful, since if the terms $f(\xi_j)$ in the normalized

sum $S_n(f)$ are small, then the cancellation effect of the randomization by means of the multiplying factors ε_j , i.e. the replacement of the terms $f(\xi_j)$ by $\varepsilon_j f(\xi_j)$ is negligible. This implies that the symmetrization argument is ineffective in this case.

Hence a new method had to be found to estimate the supremum of the normalized sums $S_n(f)$ if the additive terms $f(\xi_j)$ in these sums are small. This was done in paper [3]. After proving this result I recognized that it makes the symmetrization argument in the study of the original problem superfluous. Theorem 1 can be proved in a much simpler direct way with the help of the result of [3]. This was done in the present paper.

The question arose for me whether the symmetrization argument cannot be replaced by a simpler and stronger method in the investigation of other problems. In particular, it would be interesting to consider the multivariate version of the extension of Theorem 1 formulated in Theorem 8.4 of [2]. This is an estimate about the tail distribution of the supremum of appropriate degenerated U -statistics. This result was proved in [2] by means of an adaptation of the symmetrization argument. We needed a multivariate version of this method which was based on a generalized version of the symmetrization lemma presented in Lemma 15.2 of [2]. The proof of this lemma was not difficult, but in its application we have to estimate a rather complicated conditional probability (see formula (15.3) in [2]), and this made the proof of the above mentioned Theorem 8.4 rather hard. It seems very probable that one can find a much simpler proof with the help of the method of the present paper.

References.

- [1] J. Komlós, P. Major, G. Tusnády: An approximation of partial sums of independent rv.'s and the sample DF. I. Z. Wahrscheinlichkeitstheorie verw. Gebiete 32, 111–131 (1975)
- [2] P. Major: On the estimation of multiple random integrals and U -statistics *Lecture Notes in Mathematics* vol 2079 (Springer) Heidelberg New York Dordrecht London
- [3] P. Major: Sharp estimate on the supremum of a class of partial sums of small i.i.d. random variables. submitted to Electron. J. of Probab.
- [4] H. P. McKean Jr.: *Stochastic integrals*. Academic Press New York London (1969)
- [5] M. Talagrand: New concentration inequalities in product spaces. *Invent. Math.* **126**, 505–563 (1996)
- [6] M. Talagrand: *The general chaining*. Springer Monographs in Mathematics. Springer–Verlag, Berlin Heidelberg New York (2005)